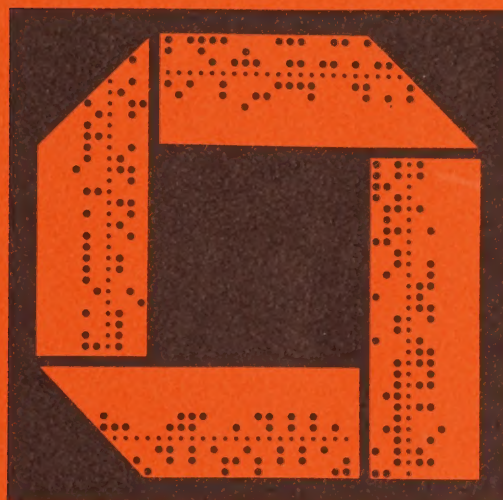


CAI  
AK  
-80M16

# THE MACHINE READABLE ARCHIVES



AN OVERVIEW  
OF ITS  
OPERATIONS AND PROCEDURES





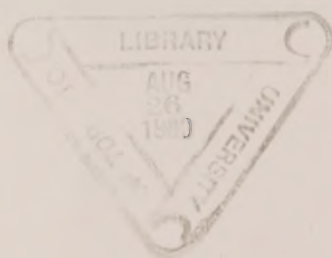
CAI  
AK  
-80411

THE MACHINE READABLE ARCHIVES:  
AN OVERVIEW OF ITS OPERATIONS  
AND PROCEDURES



Public Archives  
Canada

Archives publiques  
Canada



## Table of Contents

|  | Page |
|--|------|
| FOREWORD .....                                       | v    |
| INTRODUCTION .....                                   | 1    |
| BACKGROUND .....                                     | 2    |
| History of Electronic Data Processing .....          | 2    |
| EDP in the Government of Canada .....                | 2    |
| MACHINE READABLE ARCHIVES DIVISION .....             | 4    |
| Background .....                                     | 4    |
| Establishment of the Machine Readable Archives ..... | 4    |
| Activities .....                                     | 7    |
| APPRAISAL/ACQUISITION .....                          | 8    |
| Policy .....   | 8    |
| Appraisal .....                                      | 9    |
| Archival Considerations .....                        | 9    |
| Technical Considerations .....                       | 9    |
| Data Bases .....                                     | 10   |
| Research Contracts/Royal Commissions .....           | 10   |
| ACCESSIONING/PROCESSING .....                        | 11   |
| Procedures .....                                     | 11   |
| Cataloguing .....                                    | 12   |
| CONSERVATION .....                                   | 13   |
| SERVICE .....  | 14   |
| Access .....   | 14   |
| Confidentiality .....                                | 14   |
| CONCLUSION .....                                     | 15   |
| APPENDIX I — DEFINITIONS .....                       | 16   |
| APPENDIX II — CHART OF EDP EXPENDITURES .....        | 17   |
| BIBLIOGRAPHY .....                                   | 18   |



Digitized by the Internet Archive  
in 2022 with funding from  
University of Toronto

<https://archive.org/details/31761115507121>

# Foreword

This *Guide* is an overview of the interests, objectives, procedures and activities of the Machine Readable Archives Division, Public Archives of Canada. We hope that it will prove a satisfactory introduction to our Division, not only for archivists, but also for those outside the archival profession. We welcome your comments and inquiries.

Many thanks to those staff of the Machine Readable Archives who offered helpful comments and suggestions during the preparation of this *Guide*. Special gratitude to James H. Griffin, Head of the Computer Systems Section, for his interest and encouragement, to Monique Marsan who prepared the French text and also to Agathe Renaud and Ardyth Koeck for their patience in typing the many versions of this *Guide*.

No institution can afford to forget its founding force. The Machine Readable Archives acknowledges the efforts on its behalf of Hugh A. Taylor, former Director of the Archives Branch, and Michael E. Carroll, founding Chief of the Division.

Prepared by Nancy E. Jennings  
Social, Economic and Cultural Section  
Machine Readable Archives Division





# Introduction

The Machine Readable Archives Division was established within the Public Archives of Canada in 1973 to provide archival services for historically valuable material in machine readable form.<sup>1</sup> This development was compatible with the "Total Archives" concept described by the Dominion Archivist, W.I. Smith, as not only "...the desirability of preserving all types of archival material", but also "... the use of archival materials by or for the benefit of the greatest possible number of persons".<sup>2</sup>

The necessity of establishing a specialized division within the Archives for handling machine readable material can be viewed as a consequence of the "information revolution" of the past few decades. This

"revolution" is characteristic of any complex industrial society such as we find in Canada. As information accumulates and the awareness of it increases, manual techniques for information management and retrieval become unsuitable and machine intervention is often the only solution. This of course implies that the necessary technology is available. In the past, necessity spurred the invention of machines to deal with the problems of data processing<sup>3</sup>. Today, when we speak of electronic data processing (EDP) we are referring to data processing which is performed by electronic equipment with little human intervention. The computer, which is the heart of electronic data processing, developed according to the social needs and pressures for such a device.

<sup>1</sup> Machine readable refers to data in a physical form suitable for computer manipulation such as punched cards, paper tapes, magnetic tape, disk, drum or optic sense sheets.

<sup>2</sup> Smith, W.I., "Introduction", *ARCHIVES Mirror of Canada Past* (Toronto: Public Archives of Canada, University of Toronto Press, 1972), pp. 18 & 20.

<sup>3</sup> By data processing is meant any operation or combinations of operations performed on data (information) from the time of their creation to their destruction.

# Background

## History of Electronic Data Processing

Man has always used devices to assist him with his counting tasks. These devices have progressed through the ages from simple knots in strings to the Oriental Abacus of around 1,500 B.C. It was not until the sixteenth and seventeenth centuries that mechanical calculating devices began to appear. Some of these inventions included William Oughtred's slide rule and a logarithm gadget developed by John Napier called Napier's Bones. The most notable invention of the seventeenth century and perhaps the first precursor to the computer was Blaise Pascal's adding machine. In the eighteenth century a French weaver, Joseph Marie Jacquard, invented a system of weaving patterns in cloth through the use of a punched card.

Two remarkable inventions of the nineteenth century bring our picture of the present day computer nearer to focus. Charles Babbage's analytical engine contained all the principles of modern computers, without the electronic technology. However, his invention was forgotten only to be re-invented almost a century later when an urgent need for it arose. The first use of machine tabulation by the American government was in response to the needs of the Bureau of the Census during the 1880s. Herman Hollerith, an employee of the Bureau, developed an automated tabulating machine which operated on the principle of the punched card. The emergence of this machine allowed the Bureau to meet its deadline and marked the beginning of the era of automated data processing.<sup>4</sup>

The modern computer is the result of a series of continuous stages of development which began around 1925 with Dr. Vannevar Bush's device for solving differential equations. In 1939 Howard Aiken, in collaboration with IBM, developed the Automatic Sequence Controlled-Mark 1. This electro-mechanical device, which used telephone network relays to store and manipulate numbers, was obsolete by 1945. That year relays were replaced by vacuum tubes in the Electronic Numerical Integrator and Computer (ENIAC) developed by John Mauchly and J. Presper Eckert. The ENIAC was the first computer in the modern sense.<sup>5</sup>

With the emergence of ENIAC, computer developments increased rapidly during the 1950s. The first stored-program computer to appear was the Electronic Discrete Variable Automatic Computer (EDVAC), followed by a more functional version, the Electronic Delay Storage Automatic Calculator (EDSAC). In 1950 the first computer to use a magnetic drum, Universal Automatic Computer (UNIVAC), was built for the U.S. Bureau of the Census and continued in use until 1963. The vacuum tube was replaced by the integrated circuit for increased power, greater speed and fewer maintenance problems.

Although the history of computing devices dates back several centuries, the history of the modern computer is just over 50 years old. This may be surprising in view of the fact that computer technology is involved in so much of our everyday life. Activities in space exploration may be the most prominent example to come to mind, but even such mundane affairs as credit records, travel reservations, payrolls and bank accounts are computer controlled.

## EDP in the Government of Canada

The utilization of electronic data processing in the Canadian government began in April 1957 with the installation of a computer at No. 1 Army Pay Ledger Unit in Ottawa. This IBM 650 system was required for 16 full days each month to audit some 40,000 pay accounts. It was expected to provide a more efficient system of audit than that formerly performed by a staff of 178.

Subsequent computer installations were approved for departments involved in large projects on which significant savings could be identified. Quite often these savings were on unit record equipment and by 1962 expenditures for punched-card equipment began to decline.

By 1961 seven departments had their own computers — National Defence, the Dominion Bureau of Statistics, Comptroller of the Treasury, National Revenue — Taxation, National Research Council, Defence Research Board and Transport. The alternatives open to other departments requiring computer services were either to use the surplus computer time of these departments or to use service bureaus<sup>6</sup> outside the government.

<sup>4</sup> Thinking of a computer as an electronic calculating machine which operates according to repeatable, variable programs can help to distinguish it from other types of calculating devices.

<sup>5</sup> It may be helpful at this point to distinguish between computers and tabulating or unit record machines. Appendix I contains definitions of both these terms.

<sup>6</sup> A service bureau is a commercial establishment which provides many types of computing services to its clients.



Once the Computer Services Bureau became operational in 1970, computer use became easier for it was no longer necessary to be in possession of a computer or to go outside government circles to obtain adequate computer service. Consequently, the use of EDP spread rapidly not only among but also within departments. Both the Royal Commission on Government Organization (Glassco Commission) and the Federal Government Interdepartmental Committee on Electronic Computers identified about 22 users of EDP, including unit record equipment, in the period 1961-1962. The first Treasury Board survey of EDP in the federal government reported about 39 EDP users for 1967-1968 and by 1972 this number had increased to around 48. Coincidental with the increase in users of EDP was the increase in EDP expenditures from four million dollars in 1961 to 100 million dollars in 1972.<sup>7</sup>

Even with the budgetary controls established by Treasury Board, “. . . direct and indirect EDP expenditures [continue to] account for a steadily increasing proportion of federal government expenditures on goods and services”.<sup>8</sup> This expensive tool is, in most cases, employed by federal departments and agencies to accomplish their objectives in the most efficient manner, allowing them “. . . to support departmental programs and to improve assistance and service to the public”.<sup>9</sup> According to W. Pajor, Director, Administrative Policy Branch, Treasury Board, “. . . most departments couldn't operate without computers.”<sup>10</sup>

The main reason for the increasing dependency on EDP is simply the impossibility of processing and/or retrieving the vast amounts of information within time limitations by any other method. Also, much of the data in machine readable form is too extensive to exist in printed form. Often, if it is printed, it is only available in an aggregated form such as are most annual and sur-

vey reports. A significant proportion of the information in machine readable form is of a statistical nature, having been coded and prepared for electronic manipulation from a data gathering document such as the Census Questionnaire, the Unemployment Insurance application form or a questionnaire designed for a particular survey.

The Public Archives of Canada (PAC) has found that EDP is extremely relevant to the needs of an archives both in the area of documentation and in the provision of access to information not in machine readable form. The National Map Collection's Automated System — MARC — MAP (*Machine Readable Cataloguing*) is perhaps the most notable example of an automated retrieval system within the PAC. The main objective of this system is to facilitate the most tedious and time-consuming part of research work, “the data - gathering or what is and where stage”. Also within the PAC, automated retrieval systems or finding aids have been developed for large collections, including the papers of some of the Prime Ministers and the records of some government departments. The Union List of Manuscripts (ULM), a listing of approximately 26,000 unpublished Canadian manuscripts held by 171 archival repositories across Canada, was prepared for computerization by the staff of the PAC Union List of Manuscripts Unit. Archivists and researchers appreciate the convenience which the automated ULM provides, especially its index-by-repository. Some archivists are advocating on-line<sup>11</sup> networks of information retrieval systems between archival institutions such as already exist between libraries. This is just a brief outline of EDP's relevance to archives. The future possibilities of this technology in the area of information storage and retrieval are unlimited.

<sup>7</sup> See Appendix II.

<sup>8</sup> Shackleton, L.A., “EDP in the Federal Government — A Statistical Review,” *Treasury Board EDP Policy Project*, 1971, p. 7.

<sup>9</sup> Information Systems Division, Administrative Policy Branch, Treasury Board, *Review of EDP in the Government of Canada 1976-1977*, 1977, p. 13. See pp. 13-19 for descriptions of the ways EDP is utilized in five specific departments.

<sup>10</sup> “Questions and Answers on Ottawa's Use of Computers,” *Canadian Data-systems*, Vol. 9, No. 9 (August, 1977), p. 41.

<sup>11</sup> On line refers to equipment capable of interacting directly with a computer. With an on-line system there are no lengthy delays in order to make changes or corrections to programs, etc. A terminal such as the one utilized by the MRA (see photo p. 5) is one example of an on-line device.

# Machine Readable Archives Division

## Background

In 1973 the Public Archives approached Treasury Board for the resources to permit the acquisition, care and servicing of historically valuable material in machine readable form. This request was based upon the precepts of existing legislation. The Public Archives Act (R.S. 1970, P. 27) entrusts the Dominion Archivist with the care, custody and control of all public records, documents or other historical material of any kind, nature or description. The Public Records Order (P.C. 1966-1749) is more specific and contains a definition of the term "public records" which includes "tapes" and "computer cards".

The Federal Government Interdepartmental Committee On Electronic Computers, which had been established by Treasury Board to monitor the use of EDP within the government, had already given attention to government EDP records and had recognized the need for an organization such as the Archives proposed. The submission which the PAC presented to Treasury Board in 1973 documented the expenditures of the Federal Government for EDP and referred to the fact that no activity had begun to schedule<sup>12</sup> or preserve the material produced by this technology. The submission included a scheduling program for EDP records and recommended that it be implemented with urgency beginning with an inventory of the EDP records of all government departments and agencies. It also justified the need to extend such attention to EDP material of national significance produced by the private sector. Until then, only three Canadian universities, all with limited funds, and a few

American institutions were involved with the preservation of Canadian machine readable data.

## Establishment of the Machine Readable Archives

Treasury Board gave approval for a somewhat lesser program to ensure that the "scheduling" and the "safe-keeping" of EDP records would be accomplished. Thus the Machine Readable Archives (MRA) was established in the fall of 1973 with a staff of two. Further approval by Treasury Board of both human and financial resources have allowed the MRA to expand its activities to include the processing and servicing of its holdings.

Presently the Division is divided into three sections — Social, Economic and Cultural; State, Science and Natural Resources; and Documentation and Public Service. In keeping with the nature of the informational content of machine readable material, archivists have been selected from a variety of disciplines, including history. Technical advice and assistance are available from departmental EDP staff who keep the MRA abreast of any changes or developments in the technology.

Since the Public Archives has not as yet acquired its own computer installation, the MRA makes use of an outside service bureau for all its computing activities. Formerly it was necessary for the archivist to be physically present at this centre in order to perform these activities. With the installation of a terminal at the MRA this is no longer necessary and computing jobs can be submitted to the computer from the Division.

<sup>12</sup> A schedule is a time table which plans the life cycle of a record from the time of its creation to its final disposition.





Archivist using terminal to process file



Tape drives at service bureau

Most of the machine readable data (files) acquired by the MRA are encoded on magnetic tape<sup>13</sup>. Data which are accepted on punched cards<sup>14</sup> are transferred to magnetic tape upon acquisition. Two tape copies of each machine readable (m/r) file are made and maintained according to specific standards<sup>15</sup>. One copy is housed at an out-of-town government location while the other remains in the Divisional tape library. This library was

<sup>13</sup> See Appendix I

<sup>14</sup> *Ibid*

<sup>15</sup> Details of these standards are contained in the section headed Conservation.

constructed to preserve the magnetic tapes in a temperature-, humidity- and dust-controlled environment.

The MRA was not the first institution of its kind and it was able to benefit from the experiences of similar government institutions in Great Britain and in the United States. Some of the principles and practices presently followed by the MRA in the areas of appraisal, acquisition, processing and service are based on the experiences of these institutions, but to a great extent the MRA has developed its own procedures.



MRA tape library

## Activities

Data archives are a very recent development. For this reason, procedures and practices are still in evolution. Organizations such as the International Association for Social Science Information Service and Technology (IASSIST) are beginning to appear. The main objectives of this organization are:

1. To encourage and support the establishment at local and national levels of information centres for data base reference, maintenance and dissemination;
2. To foster international dissemination and exchange of information on significant developments in information centres for statistical and textual machine readable data bases;
3. To coordinate on an international level programs, projects and general procedural efforts which provide an international forum for the discussion of problems relating to information centres;
4. To promote the development of professional standards and encourage the establishment of training courses for data centre personnel.<sup>16</sup>

The PAC, and more specifically the MRA, has an institutional membership in this Association. Some of the staff are individual members and are active in several of

the action groups which have been “. . . organized to find solutions to specific problems and/or to develop and compile relevant materials for specific projects”.<sup>17</sup> Members of the Canadian and American action groups meet annually to work on their specific projects. The MRA is represented at these meetings.

The Division's contribution to the development of standardized procedures for machine readable files has not been limited to the activities of IASSIST. It has been well represented on two sub-committees of the Government EDP Standards Committee with two members of the Division serving as chairman and committee member respectively. The sub-committees were established within the federal government to develop standards for the documenting of m/r files and the preparation and storage of tape for extended periods. The MRA is also participating in the work of the Automated Records and Techniques Committee of the Society of American Archivists and the International Council on Archives' Committee on Automation.

When human and financial resources permit, the Division assists other repositories in establishing machine readable archival programs. Such assistance has already been given to the Quebec National Archives, as well as to the Ontario and Nova Scotia Provincial Archives in the form of technical advice and suggestions.

<sup>16</sup> IASSIST, *Newsletter*, Vol. 1, No. 1 (November, 1976), p. 5.

<sup>17</sup> *Ibid.*



# Appraisal / Acquisition

## Policy

The first priority of the MRA's acquisition policy is to acquire for preservation those machine readable records created by government departments and agencies which have historical/research value. A second and somewhat lesser priority is the acquisition of historically/nationally valuable or significant material created within the private sector. This area has not as yet been explored to any significant degree because all available resources are required to attend to the first priority. Rather than actively pursuing interests in the private sector, a somewhat passive approach to acquisition in this area has been followed. The Division will become actively involved if, during the acquisition contacts made by other PAC divisions with the private sector, machine readable material is discovered.

It is difficult to provide an accurate figure for the number of m/r files within the federal government since it changes from day to day as files are either deleted or created. However, as a starting point, an inventory of the m/r files of government departments and agencies was conducted in 1976-1977 by the Records Management Branch of the PAC. A form was distributed to all government departments as defined in Schedules A and B of the Financial Administration Act. This was accompanied by a sheet of detailed instructions which required that a form be completed and returned for each m/r file, with the exception of working or back-up files.

From the response to the inventory it was possible to assemble a minimum figure concerning the volume of m/r files within the federal government. This figure was around 3,000 but it represented only the m/r files which had been identified by responding departments and regional offices. It is safe to assume from experience that at least ten per cent of m/r files will be of historical/research value. The inventory set the stage for a scheduling system for m/r files and it increased awareness within departments that m/r files are legitimate government records and as such should be scheduled in the same manner as are paper records. There is an advantage in scheduling m/r files which does not apply to paper records. Once file copies have been obtained by the MRA, should the file be of historical value, the magnetic tape(s) can be released for reuse as required.

A scheduling system is extremely important for m/r files since the particular features of the medium render

it crucial that the files be acquired for archival purposes as close to creation as possible. There are several reasons for this: the vulnerability of magnetic tape to destruction; the risk of losing the vital paper documentation<sup>18</sup>; the possibility of the file's creator(s) becoming unavailable over time should documentation prove inadequate; the rapidity of EDP advancement perhaps precluding the converting of obsolete material into current technology. Also, the easy reusability of the medium, together with the cost of maintenance and storage, make temptation to erase and reuse magnetic tape a very real danger to historically valuable m/r material.

With these possibilities in mind the MRA has followed an *ad hoc* acquisition policy. That is to say, departments and agencies have been contacted concerning their m/r files and assistance has been offered to departmental records management staff in the completion of the inventory and scheduling of their m/r records. The cooperation and assistance of both the creators/users of m/r records and records managers are necessary at both the identification and acquisition stage of the archival process. The Division has acquired m/r material produced by several Royal Commissions — the Royal Commission on Bilingualism and Biculturalism, the Commission of Inquiry Into the Non-Medical Use of Drugs (Ledain Commission), the Commission of Inquiry Into the Marketing of Beef, to mention a few. Material has also been acquired from government departments and from researchers funded by government grants. The Health Promotion Directorate, National Health and Welfare, has been especially cooperative in the latter area. It is expected that in the future there will be some voluntary donations of machine readable material from the private sector, especially from the research community. Once the scheduling of m/r files becomes as routine as it is for paper records acquisition work will become easier. The problem will then be one of finding sufficient resources to handle the volume.

<sup>18</sup> By documentation is meant the details of the coding system used to convert the data to a machine readable form; an exact and detailed layout of where each item is and how it can be located on the file; as well as physical features of the file.



## Appraisal

The original PAC submission to Treasury Board referred to the research value of EDP records, not only from an historical viewpoint, but also from the economic, social, political, legal and psychological viewpoints. These references give some indication not only of the selection criteria used to evaluate m/r files for permanent retention, but also of the intellectual content of these files. Raw data on almost every major concern such as crime, disease, finance, energy and the environment are found in machine readable form. As with paper records, the appraisal of m/r records for permanent retention by the Public Archives is done by the archivist. Machine readable records are appraised according to their legal, informational and evidential value using the same precepts as apply to textual records. However, there are some very significant appraisal considerations which apply only to m/r records. Basically, these considerations can be divided into two categories or levels — archival and technical. Archival considerations involve the content or “why” and “what” material should be preserved. Technical considerations, which are not so directly an appraisal feature, deal with the possibilities of preserving and/or utilizing a m/r record.

## Archival Considerations

Aside from the most obvious physical differences between other media and m/r records (m/r records would never be considered of archival value because of age or physical uniqueness), there is a distinction which is the *raison d'être* of m/r records. As mentioned previously, data exists in machine readable form because of the volume, ease and speed with which repetitive data can be electronically manipulated. In contrast, paper records which are considered too routine, repetitive or voluminous are either sampled or rejected for archival purposes not only because of the prohibitive cost of storage but also because of the impracticality of analyzing such material. Automation has provided the means for a quantitative<sup>19</sup> approach to research and now “. . . more and more routine information”<sup>20</sup> is becoming useful. Rather than merely documenting historical/social facts, m/r records are used to demonstrate the interrelationships between people and/or events.

This feature of m/r material directs us to a major concern of appraisal which is the question of how to itemize selection criteria to specific areas in the light of such overwhelming possibilities for research. It is hoped that technology will develop in such a way that the pres-

ent broad approach to selection can be maintained. The current situation allows all m/r files to be considered for permanent retention with the exception of most house-keeping records or those administrative records which are obviously of little or no research value. In the future, given the present state of technology, financial or volume considerations could dictate that the selection process follow much narrower guidelines than is presently the case.

Other archival considerations include the level of aggregation of a m/r file and the possibilities of linkage with other data files. Data in m/r form are often referred to as microdata meaning that the data are represented at the lowest reporting unit possible. The less aggregated (summarized) the data the more valuable they are for research/archival purposes. For example, data available on individuals have more potential for analysis and linkage with other data files than have such data aggregated to the group level. The utilization of such common variables<sup>21</sup> in data files as sex, age, marital status, occupation, education, etc., suggests exciting and interesting possibilities for data linkage.

## Technical Considerations

Even though a m/r file may be suitable for permanent retention according to what has been discussed under archival considerations, it may be rejected for any of the following technical reasons. The file may have insufficient documentation and consequently its utilization is either impossible or unsatisfactory. It may have become obsolete, been damaged, or have deteriorated to the point that access to its contents cannot be gained. The data may be extremely invalid according to the documentation provided with no possibility of assistance from the creator towards correcting the errors. The financial burden of processing and/or maintaining the file may be too great, as is the case with large data collections or some data bases as discussed below.

Any appraisal criteria for machine readable records must take into account all of the above archival considerations, as well as those relevant and applied to conventional records. A certain insight is also required as far as future research trends and interests are concerned. At the same time, these criteria must not be so inflexible that they eliminate the element of personal judgement from the selection process. The formulation of such criteria for m/r records is not an easy task but one which the MRA has undertaken with optimism.

<sup>19</sup>Quantitative analysis refers to the trend to substitute measurement for subjective impressions of magnitude.

<sup>20</sup>Atherton, Jay, *Appraisal of Government Records*, a paper prepared for the PAC Archives Course, 1971. Revised, 1973, p. 1.

<sup>21</sup>By variable is meant any characteristic or property which has two or more possible values into which a case could be classified. For example, sex is a variable and a case could be either male or female.

## Data Bases<sup>22</sup>

An area of considerable concern to the MRA is what to do about the vast amounts of historically valuable material held in a data base environment. Some of this material is lost by deletion as the data bases are kept current. Data bases are dependent upon sophisticated software (programs, routines, etc.) in order to be accessed for input or output. Loss or obsolescence of this software eliminates access to the information contained in a data base. Financial considerations, both in the use and maintenance of a data base, render it an impractical acquisition for archival purposes.

The only practical solution open to the MRA at this time is to acquire the historically valuable material from a data base in a form completely independent of software. The Division has acquired information from a number of data bases in this manner and is negotiating the acquisition of several others.

## Research Contracts/Royal Commissions

There has been to date little compulsion on those receiving public funds for research purposes or those involved in Royal Commission Inquiries to supply the MRA with copies of any machine readable data of archival value.

The Division has forwarded proposals to the appropriate authorities regarding this area of concern. It maintains that copies of any m/r records created at public expense should be sent to the Archives when the research/Commission is completed. This should be included as a provision in a research contract and as part of a Commission's terms of reference.

<sup>22</sup>A data base is a complex arrangement of data which is dependent upon a sophisticated set of software known as a data base management system.

# Accessioning / Processing<sup>23</sup>

During the negotiations between a donor and the MRA concerning acquisition of a m/r file, the donor is consulted regarding any restrictions on the distribution of data which is of a confidential nature. The time period of any restriction is also specified at this time. An arrangement concerning the physical transfer of the file is then made. The donor may wish to make a copy of the file for the MRA and be reimbursed for the costs incurred. It is usually the case that the file is turned over to the MRA for copying and then returned to the donor. Since the concept of an original document is not relevant to data in machine readable form, each complete copy of a file is as good as the next as far as the information it contains.

## Procedures

Once a m/r file has been physically acquired by the MRA it is assigned an accession-control number. These eight digit numbers are assigned to files in sequential order as they are acquired. This number acts as an internal control number during processing and provides the key for locating the file in the tape library as well as the file's paper documentation which is filed separately. It becomes the data set name (DSN)<sup>24</sup> of the file on the internal label<sup>25</sup> of the magnetic tape.

The file is then copied and converted to a standard physical format on a MRA tape(s) which has been assigned an external or volume serial number (vol ser). This number, as well as the DSN, is included in the tape label during the copying process. The vol ser number identifies the tape for selection from the service bureau or computer installation tape library whenever action is required on the tape. The second or back-up copy of the file may or may not be made at this time depending upon the particular circumstances.

The archivist then familiarizes himself/herself with the file by examining the documentation to ascertain if all necessary information is available. Two forms which serve as a checklist are completed and an attempt is

made to obtain any missing information. These forms concern the physical attributes of the file — track, density, etc. — and the intellectual properties such as the methodology and sampling used in the data collection.

The file information is then verified by obtaining frequencies<sup>26</sup> on each variable within the file. This is usually accomplished by use of a statistical package<sup>27</sup> such as SPSS<sup>28</sup>. Any discrepancies between the file documentation and the frequencies are noted and the file documentation is checked again. If a problem still remains unsolved, the donor is then contacted for assistance. If it is impossible to obtain an explanation of the discrepancies, they are noted as unknown errors in the codebook documentation prepared by the archivist.

Preparation of MRA documentation for each file is the next step. This documentation usually contains the following information but may vary according to the nature of the file:

- (a) Title and name of principal investigators;
- (b) Source: where the data set came from;
- (c) Date: date on which the study was transmitted to the Machine Readable Archives;
- (d) Study description:
  - 1. date of the study
  - 2. basic record unit (individuals, newspapers)
  - 3. geographical area
  - 4. language of the study
  - 5. aim of the study
  - 6. publications, if any, resulting from the study;
- (e) Methodology:
  - 1. instrument of data collection (questionnaire, coding sheet, optic sense card, etc.)
  - 2. who collected the data (interviewer, mail-out questionnaire)
  - 3. how data were collected
  - 4. how data were made machine readable (i.e., optic sense sheets, key-punched);

<sup>23</sup>Processing refers to all activities in connection with a file from the time agreement is reached to acquire it until the completion of editing of the final version of the documentation prepared by the archivist.

<sup>24</sup>Data set name — an identifier for a collection of similar and related data records in machine readable form (example — G00-00000).

<sup>25</sup>Tape label is a machine readable identifier which defines the characteristics of a file.

<sup>26</sup>Frequencies is a statistical procedure which examines and displays the distribution or occurrence of each variable and its values.

<sup>27</sup>Statistical package refers to a system of computer programs designed for the analysis of data.

<sup>28</sup>SPSS — Statistical Package for the Social Sciences.



- (f) Sample:
  1. how the basic record unit was chosen (at random, etc.)
  2. number in sample
  3. universe from which sample was chosen
  4. how sample was chosen
  5. weighting;
- (g) Processing Information:
  1. form in which data were received (tape, cards)
  2. condition of data
  3. particular problems with regard to the processing of the data should be noted;
- (h) Codebook: All information on the location and description of variables should be explained in as much detail as possible;
- (i) Appendix:
  1. a copy of the original questionnaire(s)
  2. any information deemed necessary to give further explanation of variables
  3. diagrams, etc.

The final activity of the processing procedure includes the completion of a pre-cataloguing worksheet and a file log sheet. The former contains all the information required to produce a catalogue entry for the file. The log sheet contains all pertinent information about the file, including the date each copy was made and where each is located. When the archival process is completed, the final version of the documentation is referred to a senior archivist for review. All other information — problems encountered in processing, final copy of frequency and/or other computer runs, the date the file was acquired and processed, all pertinent documents relating to the file's acquisition, that is, letters of authorization or other correspondence — is included with the documentation of the file. Two fresh archival file copies should be made at this final stage of the processing.

This description of the processing procedures for a m/r file of long term value has not included the multitude of problems which are, quite frequently, encountered at this stage of the archival process. These problems, which are usually the result of inadequate documentation or are of a technical nature, account for a major portion of the time associated with file processing.

## Cataloguing

The MRA is implementing the Anglo-American Cataloguing Rules (AACR-II) as revised for machine read-

able files by the American Library Association Subcommittee on Rules for Cataloguing Machine Readable Data Files (MRDF)<sup>29</sup>. Since the AACR are flexible enough to be applied to the description of machine readable files, it seemed advisable to use a familiar system of cataloguing rather than introduce a new system which might create problems for both researchers and cataloguers alike.

The exceptions to be made in the cataloguing of MRDF represent the physical differences between this medium and that of printed material since the intellectual content would be identical in either form. For example, the requirements of the medium are such that creation of a completely detailed description of a MRDF would be impractical since the extensive catalogue records produced would tend to confuse rather than assist the user. Also, the ready convertibility of a MRDF to a different form or format would render any catalogue entry containing such information inaccurate or extremely costly to maintain.

For these reasons a multi-level system of documentation is desirable. Such a system involves four distinct types of records: standard catalogue entries (see example); data abstracts (data description forms); content documentation or codebook; and a record of the physical and logical characteristics of the file. These records would be interrelated as are the complex records necessary for the utilization of other types of exceptional library materials. The MRA is presently involved with these four levels of documentation. It is also investigating the possibility of providing microfiche copies of its catalogue holdings to libraries and related institutions.

### EXAMPLE: CATALOGUE CARD FOR MACHINE READABLE DATA FILE

MRA Canada. Royal Commission on Bilingualism and G00- Biculturalism.  
00001

Young people's images of Canadian society [machine readable data file]: an opinion survey of Canadian youth 13 to 20 years of age: NORC youth survey / John W.C. Johnstone [researcher for the Commission]. — 2nd ed. — Toronto: Institute for Behavioral Research; [Ottawa: also distributed by Public Archives Canada, Machine Readable Archives] 1973.

1365 logical records.

Bilingual codebook available.

Related to Studies of the Royal Commission on Bilingualism and Biculturalism, no. 2 with the same author and title. Ottawa, Queen's Printer, 1969.

Summary: Views of 1,365 young Canadians during May to July 1965 on the nature of their country, its people and its problems, and on their own expectations about living in Canada as adults. Social issues and in particular the question of bilingualism in the nation has been probed.

I. Johnstone, John W.C. II. Institute for Behavioral Research. III. Title. IV. Title: NORC youth survey.

<sup>29</sup> Machine Readable Data File was the term accepted by the ALA Subcommittee as the standard reference for machine readable files.



# Conservation

The vulnerability of magnetic tape and the need to find a more compact storage medium prompted the MRA and other divisions within the Archives Branch of the PAC to investigate possible alternatives. Because of the financial constraints associated with current technology, the decision was made to retain magnetic tape as the storage medium for the Division. A few preventive measures were introduced to relieve some of the worries concerning the vulnerability of magnetic tape until use of a more stable medium became feasible. These measures provide that two copies of each m/r file are to be maintained in separate physical locations. A Computer Output Microfilm (COM) copy of the file is also to be made and retained. The m/r tapes are stored under ideal humidity-, temperature- and dust-controlled conditions.

These ideal conditions include a tape library where temperatures are kept between 18.5 - 23.5°C, humidity is maintained between 40-60 degrees and dust is filtered to 50 microns. The library has an emergency backup power supply in order to maintain these conditions should a power failure occur.

Since uneven stress buildup can, in time, permanently damage a magnetic tape, the MRA has decided to have each file precision rewound once every year. The rewinding is done on a machine specifically designed for that purpose. This machine is located in the MRA's tape library and is used only by trained personnel. In addition, fresh copies of each file copy will be produced about every four years.



Precision rewinding of tape

# Service

One function of the MRA is the provision of advice, assistance and reference services to researchers in both government departments and the private sector. This is perhaps the most rewarding aspect of archival work and it is no less so for the MRA. The availability of m/r data within the PAC not only permits a different kind of research, it also attracts a new type of researcher. These new researchers include the political scientist, sociologist, psychologist, physical scientist and economist, as well as the historian. This community of researchers will increase steadily over the years ensuring the use of archives by and for the benefit of the greatest number of persons.<sup>30</sup>

The kinds of service which the MRA offers in relation to the machine readable material in its custody include the provision of tape copies or extracts from them in a variety of physical formats to meet the requirements of the individual researcher. The Division also performs statistical analysis of its data upon request.

The present cost of these services to researchers is a nominal charge designed to recover the cost of the tape and the processing expenses incurred in performing the particular service. An agreement is signed at the time the service is requested containing all the conditions, including the service cost, under which the service will be performed. Included in these conditions is the proviso that the copied material provided to researchers is for their exclusive use and cannot be resold or otherwise transferred.

## Access

The MRA has considered several ways of disseminating information concerning its holdings and potential services. A Bulletin listing files available to researchers was sent out to departmental libraries, records managers, universities and other archives throughout Canada and the United States. This Bulletin was updated

periodically and sent out whenever a sufficient number of new files became available. As an alternative to the Bulletin, the Division produces and disseminates a holdings catalogue at least once per year. In addition, special publications are produced periodically. The first such publication, a brochure entitled *Drug Use Files*, catalogues and summarizes each of the files in the collection pertaining to drug use.

## Confidentiality

As previously mentioned, much of the material in m/r form has been accumulated for the purpose of a survey or for other statistical purposes from a questionnaire or form. It often contains information of a personal, identifying nature. A concern has been how, and if, this material could be made available while at the same time protecting the privacy of the individual. The problem is not exclusive to the MRA, and the Division must be in a position to advise a donor concerning the alternatives to restricting completely the use of a file containing confidential information.

The MRA has accepted three alternatives as policy. First, the file may be restricted only as long as it is a threat to privacy. Secondly, confidential data can be released to certain researchers approved by the donor and/or the MRA to respect the anonymity of individuals. Thirdly, data can be released in a form such that individuals cannot be identified.

The MRA is sufficiently concerned about confidentiality that a position is to be reserved on staff for a person to specialize in the examination of microdata for research purposes according to the precepts described above. The duties of this position also include familiarization with the research community and assistance to donors concerned about the confidentiality of their data.

<sup>30</sup> See footnote 2.

# Conclusion

The preceding pages, which have described the operations of the Machine Readable Archives Division, have alluded to some major implications for archives. Basically, they are of a technical/financial nature for there can be no doubt as to the validity of computerized data as an archival concern. It does appear inevitable that EDP records will at least be considered by most archives before too long. Whether or not it will be feasible for an archives to accommodate this material will depend mainly upon its financial resources which could, in turn,

depend upon the state of the technology. If, for example, an economical method of storage is developed and if EDP costs can be greatly reduced, it may be possible for smaller archival institutions with limited budgets to maintain computerized material. Whatever the outcome, because of electronic technology there will be a more complete record of our lives available to our successors. Keeping this record is what a machine readable archives is all about.

# Appendix I

## DEFINITIONS

### Computer

A computer is a mechanism capable of manipulating data. Unlike a desk calculator, which must be manually guided step by step, a computer is capable of guiding most of its own operations. Some solid-state desk calculators, however, have started to earn the name of computers since certain functions, such as statistical calculation, have been wired-in permanently.

Basically, today's computers consist of:

1. Input devices to bring in information from the outside into the computer (for instance, punch-card readers, or magnetic-tape drives);
2. Memory devices (such as core memory, or thin-film memories) capable of storing the information before, during and after the processing of that information;
3. An arithmetic (or logic) unit which will operate (numerically or logically) on the information;
4. A control unit which interprets the instructions given to the computer, transfers the information from unit to unit and permits the arithmetic unit to perform the required operations;
5. Output devices which bring out the information in an acceptable form (such as the printer, or a cathode ray tube);
6. A repertory of conditional jump instructions, that is, instructions which will transfer control to certain instructions depending on the outcome of simple tests such as comparing two numbers;
7. The ability to store a program internally, that is, to store a sequence of elementary instructions by

means of which the computer user can have the computer carry out any information processing required.

### Magnetic Tape

Magnetic Tape is tape with a magnetic surface on which data can be stored by selective magnetization. Data are recorded on the tape by magnetizing narrow lengthwise strips (called tracks) in alternating directions. Magnetic tapes are usually described as either seven or nine track. Common densities of recorded information are 200, 556, 800, 1600 and 6250 bits-per inch (bpi). The MRA uses nine track, 1600 bpi tapes.

### Punched Cards

Punched cards feed data into the computer according to the rectangular holes punched in each card.

### Tabulating Equipment or Unit Record Equipment

Electro-mechanical machines which perform fairly limited functions. For instance, they are capable of examining (reading) punched cards and, according to breaks in the control fields, are able to make the necessary distinctions to order data which are eventually totalled and printed out. Since these data are usually printed out in table form these machines have become known as tabulators.



## Appendix II

### FEDERAL GOVERNMENT FULL EDP COSTS 1973-74 TO 1978-79

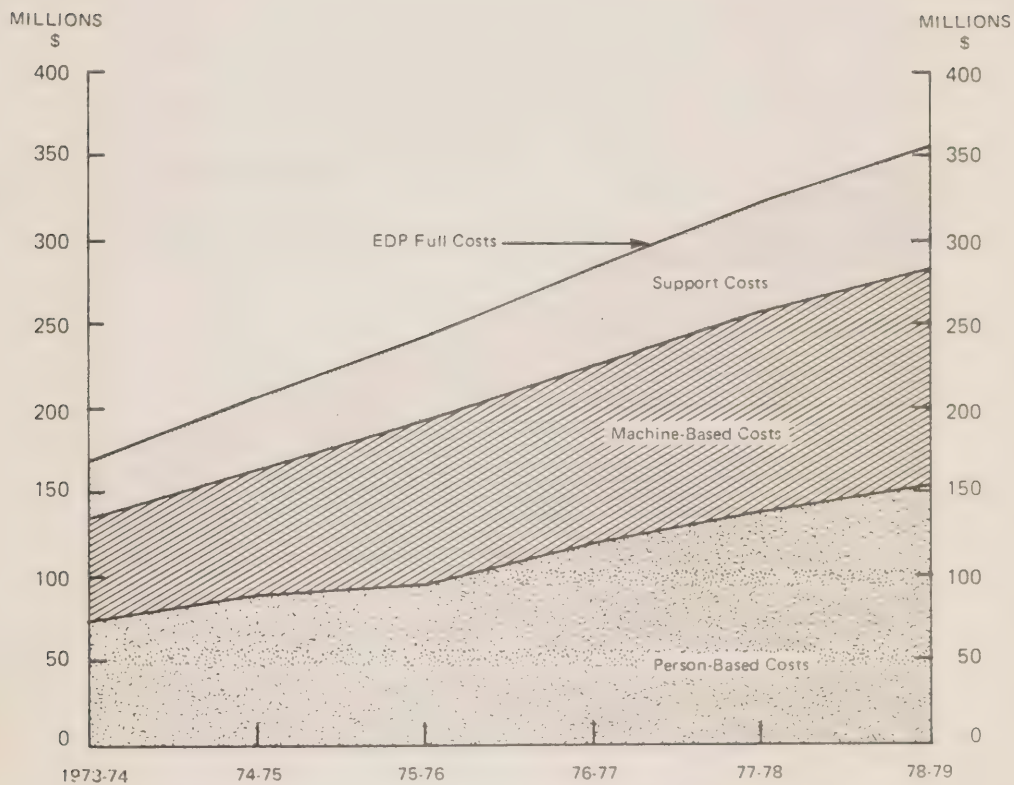


Chart taken from *Review of EDP in the Government of Canada 1976-1977*, Information Systems Division, Administrative Policy Branch, Treasury Board, 1977.

# Bibliography

- ARCHIVES Mirror of Canada Past*. Toronto: Public Archives of Canada, University of Toronto Press, 1972.
- Atherton, Jay. *Appraisal of Government Records*. A paper prepared for the Public Archives of Canada Archives Course, 1971. Revised, 1973.
- Information Systems Division, Administrative Policy Branch, Treasury Board. *Review of EDP in the Government of Canada 1976-1977*. 1977.
- International Association for Social Science Information Service and Technology (IASSIST). *Newsletter*, Vol. 1, No. 1 (November, 1976).
- "Questions and Answers on Ottawa's Use of Computers." *Canadian Datasystems*, Vol. 9, No. 9 (August, 1977).
- Shackleton, L.A. "EDP in the Federal Government — A Statistical Review." *Treasury Board EDP Policy Project*. 1971.









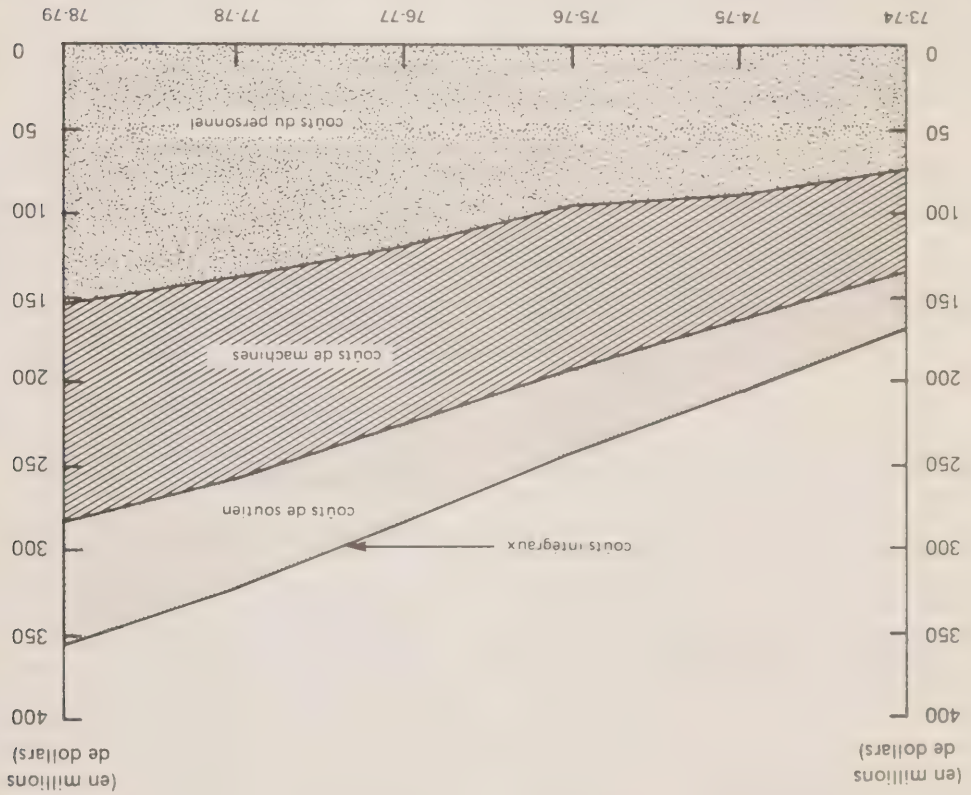
# Bibliographie

- Archives : Miroir du passé du Canada*. Archives publiques du Canada et Presses de l'université de Toronto, 1972.
- ATHERTON, Jay. *Appraisal of Government Records*. Texte préparé pour le cours d'archivistique des Archives publiques. Archives publiques du Canada, 1971; édition révisée, 1973.
- Division des systèmes d'information, Direction de la politique administrative, Conseil du Trésor. *Revue de l'information au gouvernement du Canada, 1976-1977*. 1977.
- IASSIST. *Newsletter*, vol. 1, n° 1, novembre 1976.
- Questions and Answers on Ottawa's Use of Computers*, dans *Canadian Data-systems*, vol. 9, n° 9, août 1977.
- SHACKLETON, L.A. *EDP in the Federal Government — A Statistical Review*, dans *Treasury Board EDP Policy Project*, 1971.



## Annexe II

### COUTS INTÉGRAUX DE L'INFORMATIQUE DANS L'ADMINISTRATION FÉDÉRALE DE 1973-1974 À 1978-1979



Ce tableau a été tiré de la *Revue de l'informatique au gouvernement du Canada, 1976-1977, 1977*, publiée par la Division des systèmes d'information, Direction de la politique administrative, Conseil du Trésor.

# Annexe I

## DÉFINITIONS

### Ordinateur

Un ordinateur est une machine capable de manipuler des données, mais, contrairement à un calculateur de bureau que l'on fait fonctionner manuellement étape par étape, l'ordinateur peut accomplir lui-même la plupart de ses propres opérations. Cependant, quelques calculateurs transistorsés peuvent maintenant être appelés ordinateurs depuis que certaines de leurs fonctions, comme le calcul des statistiques, ont été programmées de façon permanente.

De nos jours, les ordinateurs se composent des éléments suivants :

1. des dispositifs d'entrée qui permettent d'introduire des informations dans l'ordinateur (par exemple, des lecteurs de cartes perforées ou des dérouleurs de bandes magnétiques);
2. des dispositifs de mémoire (mémoire à tores magnétiques ou mémoire magnétique à couche mince) pouvant emmagasiner les informations avant, durant, et après leur traitement;
3. une unité arithmétique (ou logique) effectuant des opérations (de façon numérique ou logique) sur l'information;
4. une unité de contrôle, qui interprète les instructions données à l'ordinateur, fait passer l'information d'une unité à une autre et permet à l'unité arithmétique de faire les opérations requises;
5. des dispositifs de sortie qui font sortir l'information sous une forme acceptable (imprimante ou tube à rayons cathodiques);
6. un répertoire d'instructions de branchement conditionnel, c'est-à-dire des instructions qui font passer le contrôle à d'autres instructions selon le résultat de simples tests comme la comparaison de deux chiffres;

### Bande magnétique

La bande magnétique présente une surface sur laquelle on peut enregistrer des données par magnétisation sélective. L'enregistrement s'effectue en magnétisant de fines bandes (appelées pistes) sur toute la longueur de la bande, dans un sens puis dans l'autre. Les bandes magnétiques ont habituellement sept ou neuf pistes. Les densités les plus courantes de l'information enregistrée sont de 200, 556, 800, 1600 et 6250 bits au pouce (BPI). La Division des archives ordi-

nginales permet d'introduire les données dans l'ordinateur par l'intermédiaire de trous rectangulaires perforés dans chaque carte.

### Cartes perforées

Matériel de tabulation ou matériel classique à cartes perforées

Ce sont des machines électromécaniques qui accomplissent des fonctions plutôt limitées. Par exemple, elles peuvent examiner (lire) des cartes perforées et, selon les arrêts présents dans les zones de contrôle, sont en mesure de déceler les distinctions nécessaires pour trier les données qui sont éventuellement totalisées et imprimées. On appelle ces machines tabulateurs parce que les données sont habituellement imprimées sous forme de tableau.

# Conclusion

Cet ouvrage a décrit les activités de la Division des archives ordinolingues et les répercussions importantes qu'elles ont sur l'archivistique. Ces répercussions sont surtout d'ordre technique et financier car il ne fait aucun doute que les données informatives ont une valeur incontestable en archivistique. Il semble inévitable qu'avant longtemps la plupart des dépôts d'archives tiendront compte de ces documents. La possibilité, pour les dépôts d'archives, d'exploiter des systèmes informatiques dépendra surtout de leurs ressources financières

qui, à leur tour, seront en fonction des progrès technologiques. Si l'on inventait, par exemple, une méthode économique d'entreposage et si l'on réduisait considérablement les coûts informatiques, les petits dépôts d'archives à budgets restreints seraient également en mesure de conserver des données informatiques. Quoi qu'il en soit, par l'entremise de l'électronique, les générations futures connaîtront mieux la vie de leurs pères. Présenter le passé en prévision de l'avenir, telle est la raison fondamentale des archives ordinolingues.



Une des tâches de la Division des archives ordinolinguistiques consiste à produire des conseils et de l'aide aux chercheurs des secteurs public et privé, et à leur fournir des services de référence. C'est peut-être l'aspect le plus enrichissant du travail d'un archiviste et de toute la division. L'obtention de données ordinolinguistiques par les Archives publiques a non seulement donné jour à un nouveau genre de recherches mais aussi à un nouveau type de chercheur. Ces derniers proviennent de plusieurs disciplines telles que : sciences politiques, sociologie, psychologie, sciences physiques, économie et histoire. Avec les années le nombre de ces spécialistes augmentera, assurant ainsi une utilisation croissante des archives au profit d'un nombre grandissant de personnes.<sup>30</sup>

Parmi les services offerts par la Division des archives ordinolinguistiques, notons qu'elle fournit des copies de bandes magnétiques ou d'extraits de ces bandes divers formats, selon les exigences du chercheur, et qu'elle fait sur demande l'analyse statistique de ces données.

À l'heure actuelle, les chercheurs doivent payer une somme nominale pour défrayer le coût de la bande magnétique et des frais d'ordinateur. On fait alors signer une entente entre les deux parties renfermant les conditions et le coût du service. Elle stipule également que la copie du document n'est destinée qu'à l'usage exclusif du demandeur et ne peut être ni revendue ni cédée.

La Division des archives ordinolinguistiques a envisagé plusieurs façons de diffuser des informations relatives à ses archives et aux services qu'elle peut offrir. Un bulletin contenant une liste des fichiers mis à la disposition des chercheurs a été envoyé aux bibliothèques des ministères, aux gestionnaires de documents, aux universités et aux autres dépôts d'archives du Canada et des États-Unis. Ce bulletin est mis à jour périodiquement et distribué chaque fois qu'un nombre suffisant de nouveaux fichiers a été acquis. Nous songeons en outre à produire

## Accès

Comme nous l'avons déjà mentionné, la plupart des données ordinolinguistiques ont été compilées dans le cadre d'un sondage, d'une enquête ou pour toute autre fin statistique à partir d'un questionnaire ou d'un formulaire. Ces données contiennent souvent des informations personnelles ou révélatrices. Nous nous préoccupons de la possibilité et de la façon de mettre ces informations à la disposition des chercheurs tout en protégeant la vie privée des individus. C'est un problème auquel nous ne sommes pas les seuls à faire face et nous devons pouvoir proposer aux donateurs d'autres solutions que celle d'interdire totalement l'accès à un dossier contenant des documents confidentiels.

La Division des archives ordinolinguistiques a adopté trois solutions possibles. Premièrement, la consultation du fichier est permise en autant qu'elle n'empêche pas sur la vie privée d'une personne. Deuxièmement, le donateur lui-même ou la Division des archives ordinolinguistiques peut autoriser la communication de données confidentielles à certains chercheurs, lesquels sont tenus de respecter l'anonymat des individus. Troisièmement, les données peuvent être communiquées sous une forme qui ne divulguera pas l'identité des individus.

Cette question du caractère confidentiel des données a suffisamment préoccupé la Division des archives ordinolinguistiques pour qu'elle ait décidé de créer un poste dont les fonctions consistaient uniquement à examiner, d'après les principes précédents, les microdonnées qui serviront à la recherche. Le titulaire de ce poste devra connaître le monde des chercheurs et conseiller les donateurs quant au sort réservé à leurs données confidentielles.

## Caractère confidentiel des données

Comme nous l'avons déjà mentionné, la plupart des données ordinolinguistiques ont été compilées dans le cadre d'un sondage, d'une enquête ou pour toute autre fin statistique à partir d'un questionnaire ou d'un formulaire. Ces données contiennent souvent des informations personnelles ou révélatrices. Nous nous préoccupons de la possibilité et de la façon de mettre ces informations à la disposition des chercheurs tout en protégeant la vie privée des individus. C'est un problème auquel nous ne sommes pas les seuls à faire face et nous devons pouvoir proposer aux donateurs d'autres solutions que celle d'interdire totalement l'accès à un dossier contenant des documents confidentiels.

et à diffuser un catalogue des fonds au moins une fois l'an. De plus, des publications spéciales paraissent périodiquement. La première, une brochure intitulée *Dossiers sur l'usage des drogues*, est un inventaire et un résumé des dossiers de notre collection sur l'usage des drogues.

# Conservation

Vu la fragilité de la bande magnétique et la nécessité de trouver un support d'entreposage plus compact, la Division des archives ordinolingues ainsi que d'autres divisions de la Direction des archives (A.P.C.) ont cherché d'autres alternatives à ce problème. Cependant, en raison des contraintes financières de l'heure, on a décidé de garder la bande magnétique comme support d'entreposage en prenant néanmoins des mesures préventives jusqu'à ce qu'on puisse utiliser un support plus stable. Ces mesures prévoient que deux copies de chaque fichier ordinolingue seront conservées dans des endroits séparés. Une sortie d'ordinateur sur microfilm doit également être faite et gardée. Les bandes magnétiques sont entreposées dans un local climatisé et exempt de poussière.

Ces conditions idéales se retrouvent dans une bande-technique où la température est maintenue entre 18,5 et 23,5° C et l'humidité entre 40 et 60 degrés, tandis que la poussière est filtrée à 50 microns. En cas de panne de courant, une source d'alimentation électrique de secours permettra de continuer à assurer ces conditions atmosphériques. Comme des tensions inégales risquent d'endommager irrémédiablement un ruban magnétique, la Division des archives ordinolingues a décidé de rebobiner ses fichiers une fois l'an, sur une machine de précision spécialement conçue à cet effet. Elle se trouve dans la bandotèque de la division et seul un personnel qualifié est autorisé à l'utiliser. De plus, de nouvelles copies de chaque fichier seront produites environ tous les quatre ans.



Rebobinage de précision d'une bande magnétique

D.A.O. Canada. Commission royale sur le bilinguisme et le  
G00- biculturalisme

00001

La société canadienne vue par les jeunes (fichier de  
données ordinolingues) : un sondage d'opinion de la  
jeunesse canadienne de 13 à 20 ans : sondage NORC/  
John W.C. Johnstone (chercheur de la commission).  
— 2<sup>e</sup> éd. — Toronto : *Institute for Behavioral Re-  
search*; [Ottawa : également distribué par les Ar-  
chives publiques du Canada, Archives ordinolin-  
gues], 1973.

1 365 enregistrements logiques.

Résumé : Opinions de 1 365 jeunes Canadiens, re-  
cueillies de mai à juillet 1965, sur leur pays, ses habi-  
tants et ses problèmes, ainsi que sur la façon dont ils  
envisagent leur vie d'adultes au Canada. Le sondage  
a porté sur des questions sociales, particulièrement  
sur le bilinguisme au pays.  
I. Johnstone, John W.C. II. *Institute for Behavioral  
Research*. III. Titre. IV. Titre : *NORC youth survey*.  
1969.



2. responsable de la collecte des données (interviewer, questionnaire distribué par la poste)

3. façon dont les données ont été recueillies

4. façon dont les données ont été converties sous forme ordinaire (p. ex. : feuillet à lecture optique, cartes perforées)

1. façon dont l'objet de l'étude a été choisi (au hasard, etc.)

2. nombre dans l'échantillon-nage

3. milieu où l'échantillon a été prélevé

4. façon dont l'échantillon a été choisi

5. pondération

1. forme sous laquelle les données ont été reçues (bandes, cartes)

2. état des données

3. problèmes particuliers relatifs au traitement des données

h) Dictionnaire chiffré :

tous les renseignements sur la localisation et la description des variables doivent y figurer de manière aussi détaillée que possible

1. une copie du ou des questionnaires originaux

2. tout renseignement jugé nécessaire pour définir les variables

3. diagrammes, etc.

La dernière étape du traitement consiste à préparer un bordereau de précatalogage et un registre. Le premier renferme tous les renseignements nécessaires pour effectuer une entrée au catalogue du fichier. Le second contient tous les renseignements essentiels sur le fichier, y compris l'emplacement de chaque copie et la date à laquelle elle a été faite. Une fois ce processus d'archivage terminé, la version finale de la documentation est remise à un archiviste d'expérience pour fin de révision. Tous les renseignements suivants font également partie de cette documentation : problèmes survenus lors du traitement; version finale des sorties d'ordinateur, soit les calculs des fréquences des variables, et de tout autre déroulement de traitement essentiel; date d'acquisition et de traitement du fichier; ensemble des documents pertinents ayant trait à l'acquisition, c'est-à-dire les lettres d'autorisation et toute autre correspondance. Deux

Catalogage

nouvelles copies du fichier d'archives complètent l'opération.

Cette description des procédés de traitement d'un fichier ordinaire d'une valeur permanente ne tient pas compte de la multitude de problèmes qui surgissent fréquemment à ce stade du processus d'archivage. Ces problèmes, qui sont d'ordre technique ou proviennent habituellement d'un manque de documentation, occupent la plus grande partie du temps consacré au traitement d'un fichier.

La Division des archives ordinaires applique les Règles de catalogage anglo-américaines (R.C.A.A.-11), révisées en fonction des fichiers ordinaires par le Sous-comité des règles de catalogage des fichiers de données ordinaires (MRDF)<sup>29</sup> de l'*American Library Association* (ALA). Puisque ces règles sont suffisamment souples pour être appliquées à la description des fichiers ordinaires, la Division des archives ordinaires a jugé bon d'utiliser un système déjà connu plutôt que d'introduire un nouveau système susceptible de créer des problèmes aussi bien pour les chercheurs que pour les catalogues.

Les restrictions à apporter dans le catalogage des fichiers ordinaires correspondent aux différences matérielles existant entre ce support et celui des textes imprimés puisés, sous une forme ou une autre, la teneur intrinsèque en serait identique. Par exemple, les caractéristiques propres de ce médium sont telles qu'il ne serait pas pratique de décrire en détail un fichier ordinaire à cause du grand nombre de notices catalographiques ainsi créées qui ne feraient qu'embrouiller le lecteur plutôt que l'aider. De plus, il est tellement facile de changer la forme d'un fichier ordinaire que la tenue à jour d'un catalogue détaillé serait imprécise ou très onéreuse.

Tout cela milite en faveur d'un système de documentation multiple qui comprendrait quatre types de documents : les fiches catalographiques normalisées (voir l'exemple), les résumés de données (formulaires de description des données), la documentation du contenu ou le dictionnaire chiffré, et un relevé des caractéristiques matérielles et logiques du fichier. Ces documents sont reliés entre eux comme le sont les fiches complexes de documents de bibliothèques à teneur exceptionnelle, toutes munies d'un système de renvoi. En plus de s'occuper de ces quatre niveaux de documentation, la Division des archives ordinaires étudie la possibilité de fournir son catalogue sur microfiches aux bibliothèques

<sup>29</sup>Machine Readable Data File : terme accepté par le sous-comité de l'ALA pour désigner les fichiers ordinaires.

# Enregistrement et traitement<sup>23</sup> des acquisitions

L'archiviste se familiarise alors avec le fichier en examinant la documentation pour s'assurer qu'elle est complète. Il remplit des formulaires servant de liste de contrôle et s'efforce d'obtenir les renseignements qui pourront manquer. Ces formulaires traitent des aspects matériels du fichier (ses pistes, sa densité, etc.), ainsi que des qualités intellectuelles telles que la méthodologie et l'échantillonnage utilisés pour la collecte des données.

L'information contenue dans le fichier est ensuite vérifiée après un calcul des fréquences<sup>26</sup> de chaque variable du fichier. Cela se fait généralement avec un paquet de programmes statistiques<sup>27</sup> tel que le SPSS<sup>28</sup>. L'archiviste prend note de toute divergence entre la documentation et les fréquences et révise la documentation. S'il ne peut résoudre le problème, il demande l'aide du donateur. Si ce dernier ne peut, à son tour, résoudre ou expliquer les erreurs, l'archiviste les inscrit comme des erreurs inconnues dans le dictionnaire chiffré (livre de codage) qui accompagne la documentation.

L'étape suivante consiste à préparer la documentation de chaque fichier d'après les directives de la Division des archives ordinolingues. Cette documentation contient habituellement les renseignements suivants, mais ceux-ci peuvent varier selon la nature du fichier :

- a) Titre et noms des principaux enquêteurs
- b) Source : provenance des données
- c) Date : date à laquelle l'étude a été transmise à la Division des archives ordinolingues
- d) Description de l'étude :
  - 1. date de l'étude
  - 2. objet de l'étude (individus, journaux)
  - 3. région géographique
  - 4. langue de l'étude
  - 5. but de l'étude
  - 6. publications engendrées par l'étude, s'il y a lieu

- e) Méthodologie :
  - 1. outil de collecte des données (questionnaire, feuille de codage, feuille à lecture optique, etc.)

<sup>26</sup>Les fréquences sont un procédé statistique permettant d'étudier et de résumer la répartition ou la manifestation de chaque variable et ses valeurs.  
<sup>27</sup>Un paquet de programmes statistiques est un système de programmes informatiques conçu pour l'analyse des données.  
<sup>28</sup>SPSS (Statistical Package for the Social Sciences) — Paquet de programmes statistiques pour les sciences sociales.

Au cours des négociations entre un donateur et la Division des archives ordinolingues pour l'obtention d'un fichier ordinolingue, le donateur est invité à présenter toutes les restrictions sur la publication de données jugées confidentielles, et à déterminer la durée de ces restrictions. On prend ensuite des dispositions pour le transfert du fichier. La Division des archives ordinolingues, dès réception du fichier, s'occupe d'en faire une copie et de renvoyer l'original au donateur, mais il arrive aussi à l'occasion qu'un donateur préfère nous soumettre une copie conforme de son fichier, quitte à se faire rembourser par la suite. Comme la notion de « document original » n'existe pas lorsqu'on parle de données ordinolingues, les copies d'un dossier ont toutes la même valeur informative.

## Méthodes

La Division des archives ordinolingues attribue un numéro d'acquisition de contrôle à chaque fichier ordinolingue qu'elle acquiert. Il s'agit d'un numéro de huit chiffres attribué par ordre séquentiel ; il sert de numéro de contrôle interne lors du traitement et permet de localiser le fichier dans la bandodthèque. Il sert également à identifier la documentation sur papier qui se rattache au fichier et qui est classée séparément. Ce numéro devient le nom de l'ensemble de données<sup>24</sup> du fichier inscrit sur une étiquette interne<sup>25</sup> de la bande magnétique.

Le fichier est ensuite copié et converti au format normalisé sur une ou plusieurs bandes auxqueltes on a décerné un numéro d'ordre de volume qui est un numéro externe à la bande. Ce numéro ainsi que le nom de l'ensemble de données sont inclus dans l'étiquette de la bande magnétique lors du processus de reproduction. Le numéro d'ordre de volume identifie la bande de manière à la retrouver dans la bandodthèque du centre de traitement ou du service informatique chaque fois qu'on veut s'en servir. Une deuxième copie du fichier, aussi appelée fichier de réserve, peut être faite immédiatement ou plus tard, selon les circonstances.

<sup>23</sup>Par traitement, on entend toutes les activités relatives à un fichier à partir du moment où il est accepté et jusqu'à la présentation de la version finale de la documentation préparée par l'archiviste.  
<sup>24</sup>Nom de l'ensemble de données — indicatif d'une collection de données analogues et connexes enregistrées sous forme ordinolingue (exemple : C00-00000).  
<sup>25</sup>L'étiquette interne d'une bande magnétique est un numéro inscrit sous forme ordinolingue qui définit les caractéristiques d'un fichier.

## Bases de données<sup>22</sup>

Il faut également tenir compte des possibilités qui peuvent survenir dans le domaine de la recherche, et éviter d'élaborer des critères trop rigides qui ne laisseraient aucune place à la subjectivité dans le processus de sélection. La formulation de tels critères est certes difficile, mais c'est une tâche que la Division des archives ordinolingues a entreprise avec optimisme.

fre actuellement à la Division des archives ordinolingues est d'acquérir les documents qui ont une valeur historique à l'intérieur d'une base de données tout en traitant le logiciel indépendamment. De cette façon, nous avons pu acquérir les informations de certaines bases de données et nous en négocions l'acquisition de plusieurs autres.

## Contrats de recherche et commissions royales

Jusqu'à présent, il nous a été impossible d'exiger que les chercheurs subventionnés ou les membres de commissions royales d'enquête nous fournissent des copies de données ordinolingues ayant une valeur archivistique. Nous avons transmis aux autorités compétentes des propositions en ce sens, car nous sommes d'avis que des copies de tout document ordinolingue créé aux frais du public devraient être déposées aux Archives dès que la recherche ou que les travaux d'une commission sont achevés. Ceci devrait être une des modalités de tout contrat de recherche et faire partie des obligations des commissions.

Ces dernières requièrent un logiciel perfectionné (programmes, sous-programmes, etc.) pour accéder à l'information qui s'y trouve afin d'en effectuer les entrées et les sorties. La perte ou la désuétude du logiciel entrave l'accès à cette information. Parce que le coût de l'utilisation et de l'entretien d'une base est très onéreux, il conviendrait mal d'en faire une acquisition pour des fins archivistiques. La seule solution pratique qui s'of-



Dans la demande initiale présentée au Conseil du Trésor, il est question de la valeur documentaire des documents informationnels, non seulement du point de vue historique, mais aussi du point de vue économique, social, politique, juridique et psychologique. Ces références donnent des indications sur les critères d'évaluation d'un fichier ordinaire devant être conservé en permanence, ainsi que quelques précisions sur le contenu intellectuel de ces fichiers. On retrouve sous forme ordinaire des données brutes sur les sujets d'actualité, tels que le crime, la maladie, les finances, l'énergie et l'environnement. L'archiviste évalue les fichiers ordinaires et détermine s'ils doivent être conservés en permanence en tenant compte de leur valeur juridique, documentaire ou intrinsèque, comme on le ferait s'il s'agissait de documents sur papier. Il existe toutefois des facteurs d'évaluation très importants, propres aux documents ordinaires. Nous les appelons facteurs archivistiques et facteurs techniques. Les premiers portent principalement sur le contenu de ce qu'on décide de conserver, et sur les raisons d'une telle décision. Les seconds ne constituent pas un critère d'évaluation proprement dit, mais traitent des possibilités de conservation et d'utilisation de documents ordinaires.

### Facteurs archivistiques

Malgré les différences évidentes entre les documents ordinaires et les autres genres de documents (leur valeur archivistique n'étant pas uniquement fondée sur leur âge ou leur caractère unique), il en existe une qui est la raison d'être des documents ordinaires. D'une part, les données accumulées manuellement, et considérées trop routinières, répétitives ou abondantes pour être gardées aux Archives, font l'objet d'un échantillonnage ou sont supprimées. Ceci s'explique en raison du coût exorbitant qu'occasionnerait leur entreposage et aussi parce que leur analyse en serait fort complexe. D'autre part, la facilité et la rapidité remarquables avec lesquelles l'électronique réussit à manipuler nombre de données répétitives ont donné jour aux données sous forme ordinaire. L'automatisation a ouvert la porte à une nouvelle technique de recherche appelée analyse quantitative<sup>19</sup>, avec la résultante que «...more and more routine information»<sup>20</sup> (de plus en plus d'informations ordinaires) ont acquis une nouvelle utilité. En plus de documenter des événements historiques ou sociaux, les données ordinaires servent aussi à

<sup>19</sup> La technique dite « analyse quantitative » consiste à remplacez des impressions subjectives par des mesures précises.  
<sup>20</sup> Atherton, Jay, *Appraisal of Government Records* (document préparé pour le cours d'archivistique de 1971). Archives publiques du Canada, 1971. Édition révisée.

Cette particularité des documents ordinaires nous amène à poser une question essentielle en matière d'évaluation, c'est-à-dire comment attribuer des critères de sélection précis en fonction des innombrables possibilités de recherche. Présentement, nous nous devons d'utiliser des critères d'évaluation très généraux. En effet, ces critères nous permettent de conserver en permanence tous les fichiers ordinaires à l'exception d'une grande partie des dossiers administratifs qui ne présentent aucune valeur documentaire. Il est à espérer que les progrès de la technologie feront en sorte que nous pourrions maintenir notre système actuel. Cependant, des considérations financières ou quantitatives nous poussent peut-être à adopter des critères beaucoup plus restrictifs.

Parmi les facteurs archivistiques, il faut tenir compte également de la manière dont les informations sont rassemblées dans le fichier, et des possibilités de relief des fichiers avec d'autres. Les données ordinaires sont souvent appelées micro-données, c'est-à-dire qu'elles sont réduites à leur plus simple expression. Plus les données sont détaillées, plus elles prennent une valeur documentaire et archivistique. Par exemple, il est beaucoup plus facile d'analyser des données au niveau de l'individu et de relier ces informations avec d'autres données que ce ne l'est au niveau du groupe. L'utilisation de variables<sup>21</sup> telles que le sexe, l'âge, l'état civil, l'occupation, les études, etc., dans certains fichiers offre de très intéressantes possibilités de liaison entre les données.

### Facteurs techniques

Bien qu'on puisse décider de conserver un document ordinaire en permanence, après l'avoir étudié en fonction des facteurs archivistiques mentionnés précédemment, il se peut qu'il soit rejeté en raison des facteurs techniques suivants. Le fichier peut être incomplet et par conséquent ne pas se prêter à la consultation. Il peut être désuet, abîmé ou endommagé au point qu'on ne puisse le consulter. Les données qui s'y trouvent peuvent être inexacts d'après la documentation sans qu'il soit possible de retrouver leur auteur pour les mettre à jour ou les rectifier. Enfin, il se peut qu'on ne puisse assumer les coûts de traitement et d'entretien du fichier, comme c'est le cas pour certains fichiers de données très volumineux ou pour des bases de données dont nous traiterons plus loin.

Tout critère d'évaluation des documents ordinaires doit prendre en considération tous ces facteurs, de même que ceux qui s'appliquent aux documents ordi-

<sup>21</sup> Par variable, on entend tout caractère ayant deux valeurs possibles plus, dans lesquelles on pourrait classer un cas. Par exemple, le mot sexe est une variable et le cas pourrait être masculin ou féminin.



# Evaluation et acquisition

## Politique

La politique d'acquisition de la Division des archives ordinolingues vise tout d'abord à acquérir et à préserver les documents ordinolingues émanant des ministères et organismes gouvernementaux qui représentent une valeur historique ou documentaire. En second lieu, la division tente de faire l'acquisition de documents à teneur historique ou nationale provenant du secteur privé. Jusqu'à présent elle a dû concentrer ses efforts vers son but premier, car l'absence de ressources entrave sérieusement l'exploitation du secteur privé. Elle entend bien combler cette lacune si, au hasard de leurs contacts avec le secteur privé, d'autres divisions des Archives publiques découvrent l'existence de documents ordinolingues.

Vu la difficulté de déterminer le nombre exact de fichiers ordinolingues détenus par le gouvernement, à cause de la création et de l'élimination constantes de fichiers, la Direction de la gestion des documents des Archives publiques a donc décidé en 1976-1977 de procéder à un premier inventaire. À cet effet, on a distribué une circulaire à tous les ministères relevant des annexes (a) et (b) de l'article 2 de la Loi sur l'administration financière. Les notes explicatives spécifiaient qu'un formulaire devait être rempli pour chaque fichier ordinolingue, exception faite des fichiers de travail et des fichiers de réserve, et devait être ensuite retourné à la direction.

D'après les réponses recueillies auprès des ministères et des bureaux régionaux qui ont fait suite à cette demande, on a pu identifier au moins 3 000 fichiers ordinolingues au gouvernement fédéral. Il est donc permis de supposer qu'au moins dix pour cent de ces fichiers ont une valeur historique. De cette expérience découlait l'établissement de plans de conservation pour les fichiers ordinolingues et l'acceptation par les ministères que ceux-ci constituent des documents publics au même titre que les documents sur papier. Les premiers présents cependant un avantage par rapport aux seconds; en effet, lorsque la Division des archives ordinolingues a obtenu des copies de fichiers qui ont une valeur historique, que les bandes magnétiques peuvent être réutilisées, le cas échéant.

Plusieurs facteurs entrent en ligne de compte dans l'établissement d'un plan de conservation pour les documents ordinolingues. Ainsi, la fragilité des bandes

<sup>18</sup> Par documentation, on entend les explications détaillées du système de codage utilisé pour convertir les données sous forme ordinolingues; c'est un schéma exact et précis de la position de chaque article sur la bande et de la façon d'y accéder, ainsi qu'une explication des caractéristiques du fichier.

magnétiques, le danger de perdre la documentation<sup>18</sup> ainsi que le fait que les bandes magnétiques ne peuvent être réutilisées et que le coût de l'entretien et de l'entreposage des bandes est assez élevé, on est souvent tenté d'effacer et de réutiliser une bande magnétique. Cela constitue une véritable menace pour les documents ordinolingues à valeur historique.

Dans cette optique, nous avons décidé d'adopter une politique spéciale en matière d'acquisition. Nous avons ainsi aidé les gestionnaires de documents des ministères et organismes gouvernementaux à dresser un inventaire et à fixer des plans de conservation pour les documents ordinolingues. On ne saurait trop insister sur l'importance de la collaboration qui doit exister entre les créateurs/usagers des archives ordinolingues et les gestionnaires de documents dans le processus d'identification et d'acquisition des archives. Grâce à cette collaboration nous avons pu acquérir les documents ordinolingues de plusieurs commissions royales, notamment ceux de la Commission royale sur le bilinguisme et le biculturalisme, la Commission d'enquête sur l'usage non médical des drogues (Commission Ledain) et la Commission d'enquête sur la commercialisation du bœuf. Nous avons également obtenu des documents auprès de ministères et de chercheurs bénéficiaires de subventions gouvernementales. La Direction générale de la promotion de la santé, ministre de la Santé et du Bien-être social, a particulièrement fait preuve de coopération dans ce domaine. Dès que l'établissement des plans de conservation des fichiers ordinolingues deviendra chose courante, comme c'est le cas pour les documents sur papier, notre travail d'acquisition se trouvera simplifié. Il s'agira alors d'obtenir les ressources suffisantes pour s'occuper de tous ces documents. Nous espérons qu'à l'avenir le secteur privé — et particulièrement les chercheurs — nous fera don de documents ordinolingues.

L'archivage de données ordinolingues est un phénomène très récent, ce qui explique que ses procédés et ses applications soient en constante évolution. C'est ainsi qu'on pris naissance des organismes tels que l'*International Association for Social Science Information Service and Technology* (IASSIST). Parmi ses objectifs, on note :

1. encourager et promouvoir la création de centres d'information pour bases de données à l'échelle régionale et à l'échelle nationale pour fins de référence, d'entretien et de diffusion;
2. stimuler la diffusion et l'échange, à l'échelle internationale, d'informations au sujet des progrès importants réalisés dans les centres d'information touchant les bases de données ordinolingues statistiques et textuelles;
3. coordonner, sur le plan international, programmes, projets et expériences tout en favorisant la communication entre pays sur les problèmes relatifs aux centres d'information;
4. favoriser la création de normes professionnelles et l'établissement de cours de formation pour le personnel de ces centres.<sup>16</sup>

Les Archives publiques, et plus précisément la Division des archives ordinolingues, sont membres de cette association. Certains membres du personnel y adhèrent à titre individuel et font partie de plusieurs groupes d'action « ...organized to find solutions to specific

problems and/or to develop and compile relevant materials for specific projects »<sup>17</sup> (mis sur pied pour résoudre des difficultés particulières et pour produire et compiler des documents relatifs à des projets spéciaux). Les membres de groupes d'action canadiens et américains se réunissent une fois l'an pour travailler à leurs projets spéciaux. La Division des archives ordinolingues est représentée à ces réunions.

Notre contribution à l'élaboration de procédés normaux portant sur les fichiers ordinolingues ne se limite pas aux activités de l'IASSIST. Deux membres de la division nous ont bien représentés, l'un comme président, l'autre comme membre de comité, au sein de deux sous-comités relevant du Comité de normalisation internationale du gouvernement. Les buts de ces comités étaient d'employer au gouvernement fédéral des normes de documentation pour les fichiers ordinolingues et des bandes magnétiques pour de longues durées. La division participe également aux travaux du Comité des techniques et des documents automatisés de la Société des archivistes américains et du Conseil international des archives sur l'automatisation.

Dans la limite de nos ressources humaines et financières, nous aidons volontiers d'autres dépôts d'archives à instaurer des programmes d'archives ordinolingues. C'est ainsi que nous avons déjà prêté notre concours aux Archives nationales du Québec, au Archives provinciales de l'Ontario et à celles de la Nouvelle-Ecosse, au moyen de conseils techniques et de recommandations.

thèque de la division et l'autre dans un établissement gouvernemental à l'extérieur de la ville. La température et le degré d'humidité de cette bandorhèque sont contrôlés de manière à assurer la préservation des rubans magnétiques.

Les gouvernements de la Grande-Bretagne et des États-Unis ont servi de modèles à la Division des archives ordinolingues qui a su mettre leur expérience à profit. Elle a même adopté certains de leurs principes et pratiques en matière d'évaluation, d'acquisition, de traitement et de service, mais elle a, en grande partie, développé ses propres procédés.



Bandorhèque de la Division des archives ordinolingues

La plupart des données ordinolingues (fichiers) acquises par la Division des archives ordinolingues sont codées sur bande magnétique<sup>13</sup>. Les données reçues sous forme de cartes perforées<sup>14</sup> sont transférées sur bande magnétique dès leur acquisition. On produit deux copies de chaque fichier ordinolingue sur des bandes magnétiques qui sont ensuite entreposées selon des normes déterminées<sup>15</sup>. Une copie est conservée dans la bando-

<sup>13</sup>Voir annexe I.

<sup>15</sup>On trouvera des détails sur ces normes au chapitre intitulé « Conservation ».



Detrouleur de bande dans un centre de traitement



Archiviste en train de traiter un fichier au terminal





# La Division des archives ordinolingues

Création de la Division des archives ordinolingues

Dans cette optique, le Conseil du Trésor approuva néanmoins la création d'un programme un peu moins vaste qui assurait l'établissement des « plans de conservation »<sup>12</sup> et la « garde » des documents informatiques. C'est ainsi que la Division des archives ordinolingues (D.A.O.), ne comprenant que deux employés, fut mise sur pied à l'automne de 1973. Par la suite, le Conseil du Trésor autorisa une augmentation des ressources humaines et financières, permettant ainsi à la D.A.O. d'étendre ses activités au traitement et à l'entretien de ses fonds.

Notre division se compose actuellement de trois sections, soit : la Section des archives sociales, économiques et culturelles, la Section des archives d'État, des sciences et des ressources naturelles et la Section de la documentation et du service au public. Nos archivistes, en rapport avec la nature informationnelle des documents ordinolingues, nous provenaient de différentes formations universitaires, y compris l'histoire. Ils bénéficient de l'assistance et des conseils d'ordre technique prodigués par nos informaticiens qui se tiennent au courant des progrès et des nouvelles réalisations dans le domaine de l'informatique.

Puisque les Archives publiques ne possèdent pas en core d'ordinateur, la Division des archives ordinolingues doit utiliser les services d'un centre de traitement pour ses besoins informatiques. La présence d'un archiviste au centre de traitement n'est plus nécessaire depuis que l'on a installé un terminal sur les lieux de la division même, et qu'il est maintenant possible de communiquer directement avec l'ordinateur.

En 1973, le département des Archives publiques de-manda au Conseil du Trésor les ressources nécessaires pour acquérir, conserver et garder des documents de valeur historique sous forme ordinolingue. Cette de-mande fut formulée en fonction des lois actuelles. La Loi sur les Archives publiques (S.R.C. 1970, chap. P-27) confie les documents publics, ainsi que les pièces historiques de toute espèce, nature et description, au soin, à la garde et à la régie de l'archiviste fédéral. Le Décret sur les documents publics (C.P. 1966-1749) offre encore plus de précision et englobe les expressions « bandes magnétiques » et « cartes perforées » dans sa définition de « document public ».

Le Comité interministériel du gouvernement fédéral sur les ordinateurs électroniques, créé par le Conseil du Trésor afin de contrôler le nombre d'utilisateurs de l'infor-matique au gouvernement fédéral, a reconnu, après étude de la question des documents ordinolingues, la viabilité d'un organisme tel que proposé par les Ar-chives publiques :

Dans leur soumission au Conseil du Trésor en 1973, les Archives publiques ont fait état des dépenses énormes en informatique pour le gouvernement fédéral et du fait qu'aucun « plan de conservation »<sup>13</sup> n'avait été prévu pour préserver les documents résultant de cette technologie. On proposait un programme de conserva-tion des dossiers informatiques et on recommandait sa mise en application immédiate, en commençant par un inventaire des documents informatiques de tous les ministères et organismes gouvernementaux. On pré-conisait également l'expansion de ce programme aux documents informatiques d'importance nationale en provenance du secteur privé. Jusque-là, seulement trois universités canadiennes, toutes trois à budget restreint, et quelques établissements américains s'occupaient de la conservation des données ordinolingues canadiennes.

<sup>12</sup> Le plan de conservation est un indicateur qui planifie la vie d'un document, de sa création jusqu'à sa dernière étape.

11 L'expression « en direct » désigne la possibilité de convertir directement avec un ordinateur. Dans ce système on peut modifier ou corriger des programmes sur-le-champ. Un terminal comme celui de la Division des archives ordinaires (voir photo en page 5) est un exemple de dispositif en direct.

quantités d'informations de manière différente dans les délais prescrits a propagé le besoin grandissant de l'informaticque. De plus, la plupart des données ordinaires sont trop nombreuses pour être imprimées. Bien souvent, même si elles le sont, elles ne sont disponibles qu'en bloc, comme la plupart des rapports annuels et des enquêtes. Les données ordinaires sont, dans une large proportion, de nature statistique; elles ont été codées et préparées en vue d'une manipulation électronique, à partir de documents servant à recueillir des données, comme les questionnaires de recensement, les formulaires de demande de la Commission d'assurance-chômage ou les questionnaires conçus pour des sondages particuliers.

Le département des Archives publiques du Canada (A.P.C.) a constaté que l'informaticque correspond parfaitement aux besoins de l'archivistique favorisant la documentation ou l'accès à des informations en ce qui a trait aux fonds d'archives qui ne se trouvent pas sous forme ordinaire. Le système automatisé de la Collection nationale de cartes et plans (le MARC — MAP) constitue l'exemple le plus remarquable d'un système automatisé de retrait de l'information à l'intérieur du département des Archives publiques. Ce système a pour objectif principal de faciliter l'étape la plus fastidieuse et la plus longue du travail de recherche, qui est « la collecte des données, c'est-à-dire l'étape de la sélection des données pertinentes et de leur localisation ». Dans ce même département, on a également mis au point des systèmes de retrait automatisés ou des instruments de recherche pour de vastes collections, y compris les papiers personnels de certains premiers ministres ainsi que les dossiers de quelques ministères. Le *Catalogue collectif des manuscrits canadiens inédits* répartis dans 171 dépôts d'archives au Canada, a été préparé pour l'informaticque par le personnel de l'Unité du catalogue collectif des manuscrits, aux Archives publiques du Canada. Les archives et chercheurs trouvent ce catalogue fort utile et apprécient particulièrement son index par dépôt d'archives. Certains archivistes réclament des réseaux « en direct »<sup>11</sup> de systèmes de retrait d'informaticque entre dépôts d'archives, comme cela se fait entre bibliothèques. Ce n'est là qu'un bref exposé de la façon dont on utilise l'informaticque dans les archives. En effet, les possibilités futures de l'ordinateur dans le domaine de l'emmagasinage et du retrait de l'information paraissent illimitées.

12 SHARLETON, L.A., *EDP in the Federal Government — A Statistical Review*, *Journal of the Royal Statistical Society*, 1971, p. 13. Voir les pages 13 à 20 de ce rapport pour une description de la façon dont cinq ministères utilisent l'informaticque.

13 *Questions and Answers on Ottawa's Use of Computers*, dans *Canadian Data Systems*, vol. 9, n° 9, août 1977, p. 41.

14 Un centre de traitement est un établissement commercial qui offre à ses clients une gamme de services informatiques.

L'impossibilité de traiter ou de récupérer d'énormes ne pourrait fonctionner sans ordinateurs).  
Conseil du Trésor, « ...most departments couldn't operate without computers »<sup>10</sup> (la plupart des ministères chef de la Direction de la politique administrative, au d'aide et de services au public ».<sup>9</sup> Selon M. W. Pajor, programmes de ces derniers et à améliorer la prestation manière la plus efficace, ce qui « ...sert à appuyer les outils coûteux pour atteindre leurs objectifs de la et organismes du gouvernement fédéral utilisent cet biens et services). Dans la plupart des cas, les ministères sante des dépenses du gouvernement fédéral pour les formatique constituent toujours une proportion croissante des dépenses » (les frais directs et indirects d'inpution of federal government expenditures on good and services ») (les frais directs et indirects d'inputures continue to account for a steadily increasing le Conseil du Trésor, « ...direct and indirect EDP expenditures

Malgré les mesures de contrôle financier établies par dollars en 1972.<sup>12</sup>  
quatre millions de dollars en 1961 à cent millions de augmentation, les dépenses dans ce secteur passèrent de 1967-1968 et de 48 en 1972. Parallèlement à cette ral fit état d'environ 39 usagers de l'informaticque en le Conseil du Trésor à l'intérieur du gouvernement fédéral 1962. Le premier sondage sur l'informaticque mené par classique à cartes perforées, durant la période 1961-1962, le premier sondage sur l'informaticque mené par certains d'entre eux. La Commission royale sur l'organisation du gouvernement (la Commission Glassco) et le Comité interministériel du gouvernement fédéral sur les ordinateurs électroniques révélèrent que vingt-deux usagers avaient utilisé l'informaticque, y compris le matériel classique à cartes perforées, durant la période 1961-1962. Le premier sondage sur l'informaticque mené par certains d'entre eux. La Commission royale sur l'organisation du gouvernement (la Commission Glassco) et le Comité interministériel du gouvernement fédéral sur les ordinateurs électroniques révélèrent que vingt-deux usagers avaient utilisé l'informaticque, y compris le matériel classique à cartes perforées, durant la période 1961-1962. Le premier sondage sur l'informaticque mené par

En 1970, le Bureau des services informatiques entra en opération, ce qui facilita l'usage de l'ordinateur car il n'était alors plus nécessaire de posséder son propre ordinateur ou de recourir à des services de l'extérieur pour obtenir un bon service d'informaticque. Ceci amena une utilisation plus répandue de l'informaticque, non seulement entre les ministères mais également à l'intérieur de certains d'entre eux. La Commission royale sur l'organisation du gouvernement (la Commission Glassco) et le Comité interministériel du gouvernement fédéral sur les ordinateurs électroniques révélèrent que vingt-deux usagers avaient utilisé l'informaticque, y compris le matériel classique à cartes perforées, durant la période 1961-1962. Le premier sondage sur l'informaticque mené par

En 1961, sept ministères ou organismes avaient leur propre ordinateur : le ministère de la Défense nationale, le Bureau fédéral de la statistique, le Contrôleur du Trésor, Revenu national-impôt, le Conseil national de recherches, le Conseil de recherches pour la défense, et le ministère des Transports. Les autres ministères désireux d'utiliser les services d'un ordinateur pouvaient se servir de l'ordinateur des ministères précités, lorsqu'il n'était pas en opération, ou avoir recours aux services de centres de traitement externes.<sup>13</sup>



# Antécédents

## Historique de l'informatique

L'homme a toujours utilisé des dispositifs ou des appareils pour l'aider à calculer. Au cours des siècles, ceux-ci sont passés d'une simple série de nœuds sur une ficelle à l'abaque chinois vers l'an 1500 avant Jésus-Christ. Il faudra cependant attendre les XVI<sup>e</sup> et XVII<sup>e</sup> siècles pour voir l'apparition de calculatrices mécaniques. Parmi elles, mentionnons les ossellets de Napier et la règle à calculer d'Oughtred. C'est toutefois la machine à additionner de Blaise Pascal qui reste l'invention la plus remarquable du XVII<sup>e</sup> siècle, et est peut-être considérée comme le premier précurseur de l'ordinateur. Au XVIII<sup>e</sup>, un tisserand français, Joseph Marie Jacquard, inventa le métier à tisser qui porte son nom, et qui utilise des cartes perforées.

Deux inventions remarquables du XIX<sup>e</sup> siècle se rapprochent un peu plus de l'ordinateur tel que nous le connaissons aujourd'hui. Le moteur analytique de Charles Babbage contenait tous les principes des ordinateurs modernes, mais sans la technologie de l'électronique. Cependant, son invention tomba dans l'oubli, pour resurgir près d'un siècle plus tard, répondant à un besoin particulier. Le gouvernement américain utilisa un système de tabulation mécanique pour la première fois dans les années 1880, pour un recensement. Herman Hollerith, employé au Bureau du recensement, mit au point une machine de tabulation automatique fonctionnant selon le principe de la carte perforée. Cette machine permit au bureau de terminer son recensement dans les délais prescrits et marqua le début de l'ère du traitement automatisé des données.<sup>4</sup>

L'ordinateur moderne a été réalisé par une série de mises au point successives commencées vers 1925 alors que M. Vannavar Bush perfectionnait une machine pour résoudre des équations différentielles. En 1939, Howard Aiken, en collaboration avec IBM, conçut la Mark I, à électro-mécanique emmagasinant et manipulant des chiffres au moyen de relais de réseaux téléphoniques, mais celui-ci s'avéra désuet dès 1945. Cette même année, ces relais furent remplacés par des tubes à vide dans l'intégrateur et calculateur numérique électronique (ENIAC, *Electronic Numerical Integrator and Computer*) inventé

<sup>4</sup>Pour distinguer l'ordinateur des autres machines à calculer, il faut le considérer comme une machine électronique effectuant des calculs selon des programmes variables répétitifs.

par John Mauchly et J. Presper Eckert. L'ENIAC fut le premier ordinateur des temps modernes.<sup>5</sup>

Avec l'ENIAC, les réalisations en matière d'informatique se succédèrent rapidement durant les années 1950. Le premier ordinateur à programme enregistré a été l'ordinateur électronique automatique à variable discrète (EDVAC, *Electronic Discrete Variable Automatic Computer*), qui fut suivi d'une version plus fonctionnelle, le calculateur automatique électronique à mémoire temporaire (EDSAC, *Electronic Delay Storage Automatic Calculator*). En 1950, le premier ordinateur utilisant un tambour magnétique (UNIVAC, *Universal Automatic Computer*) a été construit pour le Bureau du recensement américain et utilisé jusqu'en 1963. Le tube à vide fut alors remplacé par le circuit intégré, ce qui accrut la puissance et la vitesse de la machine et réduisit les problèmes d'entretien.

Bien que l'histoire des machines à calculer remonte à plusieurs siècles, celle de l'ordinateur moderne ne date que d'un peu plus de cinquante ans. Cela peut paraître surprenant si l'on considère la place importante qu'occupe la technologie informatique dans notre vie quotidienne. L'exploration spatiale en est peut-être l'exemple le plus frappant, mais des choses aussi courantes que les cartes de crédit, les réservations de voyages, les feuilles de paie et les comptes bancaires sont contrôlées par ordinateur.

## L'informatique au gouvernement canadien

L'utilisation de l'informatique par le gouvernement canadien remonte à avril 1957, avec l'installation d'un ordinateur dans la première unité du service de la solde, à la Défense nationale (Ottawa). Le système IBM 650 devait fonctionner seize jours par mois pour vérifier quelque 40 000 comptes de paie et être plus efficace que les 178 fonctionnaires chargés de cette tâche.

<sup>5</sup>À ce point, il est peut-être utile d'établir une distinction entre les ordinateurs et les machines à tabulation ou de type classique. Voir l'annexe I pour la définition de ces deux termes.

# Introduction

La Division des archives ordinolingues a été créée au sein du département des Archives publiques du Canada en 1973. Elle a pour mandat de conserver les documents ordinolingues<sup>1</sup> qui ont une valeur historique. Cette responsabilité est conforme au concept des « archives totales » que l'archiviste fédéral, M. W. I. Smith, décrit comme étant « ...non seulement la conservation de documents de toutes sortes... » , mais aussi « ...l'utilisation des archives par le plus grand nombre de personnes ».<sup>2</sup>

La nécessité de créer une division spécialisée au sein des Archives pour s'occuper des documents ordinolingues est une conséquence de la « révolution de l'information » qui s'est manifestée au cours des dernières décennies. Cette « révolution » est un trait caractéristique de toute société industrielle complexe telle qu'on la

<sup>1</sup> Par document ordinolingue, on entend un document présentant des données sous une forme pouvant être utilisée par un ordinateur, comme des cartes ou des rubans perforés, des rubans magnétiques, des disques, des tambours ou des « autres formes électroniques ».

<sup>2</sup> Smith, W. I. Introduction, *ARCHIVES: Mirror du passé du Canada*, Archives publiques du Canada et Presses de l'université de Toronto, 1972, pp. 19 et 21.

<sup>1</sup> Par traitement des données, on entend toute opération ou combinaison d'opérations accomplies sur les données (ou des informations) à partir de leur création jusqu'à leur destruction.

retrouve au Canada. À mesure que l'on accumule des informations et que l'on se rend compte de leur utilité, les techniques manuelles permettant de les gérer et de les retrouver deviennent insuffisantes; la machine est alors souvent le seul moyen qui permette d'effectuer efficacement ces opérations. Pour cela, il faut, bien entendu, posséder la technologie nécessaire. Au cours du passé, la nécessité a poussé l'homme à inventer des machines capables de résoudre les problèmes du traitement des données<sup>3</sup>. De nos jours, lorsqu'on parle de traitement électronique des données ou d'informaticque, on entend une opération accomplie par un matériel presque entièrement électronique. L'ordinateur, qui s'avère l'âme de l'informaticque, a évolué suivant les besoins pressants de la société pour le développement d'un tel outil.





# Avant-propos

Dans ce document, nous offrons une vue d'ensemble des intérêts, des objectifs, du fonctionnement et des activités de la Division des archives ordinolingues, Archives publiques du Canada. Nous espérons qu'il constituera une bonne présentation de notre division, non seulement pour les archivistes, mais également pour ceux et celles qui n'ont pas de connaissances en archivistique. Nous recevrons avec plaisir tous les commentaires et demandes de renseignements qui nous seront adressés.


Nous remercions tous les employés de la Division des archives ordinolingues qui nous ont fait part de leurs commentaires et de leurs suggestions, fort utiles lors de la préparation de ce guide, et en particulier M. James H. Griffin, chef de la Section des systèmes informatiques, pour son encouragement et l'intérêt qu'il y a porté. Qu'on nous permette également d'exprimer notre reconnaissance à Monique Marsan, qui a préparé la version française, de même qu'à Agathe Renaud et Ardyth Koeck, qui ont patiemment dactylographié les nombreuses versions de cet ouvrage.

Comme aucune institution ne peut ignorer l'apport de ses pères, il nous faut souligner la contribution de M. Hugh A. Taylor, ancien directeur de la Direction des archives, et de M. Michael E. Carroll, fondateur de la Division des archives ordinolingues.

Préparé par Nancy E. Jennings  
Section des archives sociales, économiques et culturelles  
Division des archives ordinolingues



LES ARCHIVES ORDINOLINGUES :  
UNE VUE D'ENSEMBLE SUR  
SES OPÉRATIONS ET PROCÉDÉS

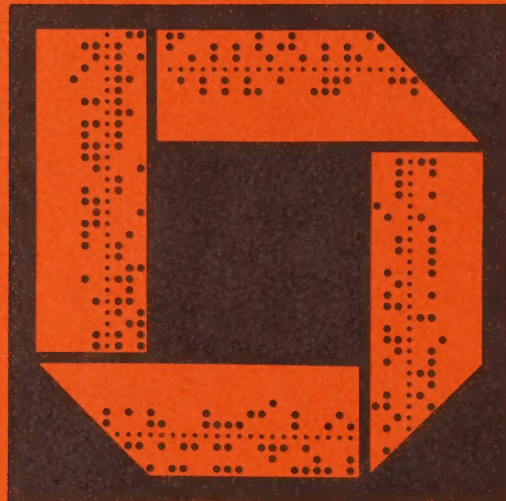
Archives publiques  
Canada  
Public Archives  
Canada







# OPERATIONS ET PROCÉDES DE TRAITEMENT DES ÉCHIFFRÉS



## LES ARCHIVES DES